

Learning-centred assessment validation framework: A theoretical exploration

Oluwaseun Ijiwade¹ and Dennis Alonzo²

¹University of New South Wales Sydney and Australasian College for Emergency Medicine, Australia

²University of New South Wales Sydney, Australia

Corresponding Author: seun.ijiwade@acem.org.au, ijiwadeseun@gmail.com; d.alonzo@unsw.edu.au

 @DrSeunIjiwade; @Dennis_A_Alonzo

(Submitted: 25 November 2022; Accepted: 26 June 2023)

Abstract

Classroom-based assessment validation has received considerable critical attention and many conceptualisations have emerged. While these conceptualisations are helpful in advancing our assessment knowledge, there is a need for a more learning-oriented teacher assessment practice validation. This paper builds on previous validation theories and approaches to redefine the validity of classroom-based assessment in terms of practical, useful, and trustworthy interpretation and uses of classroom assessment in enhancing learning and teaching. Further, the paper sets relevant inferences and prioritises teachers as sources of evidence in assessment evaluation based on pragmatic principles and Vygotsky's sociocultural theory. This explication is valuable in exploring a learning-centred validation approach for evaluating classroom assessment. The paper suggests practical principles for evaluating learning-oriented, teacher-based assessment. Lastly, the paper concludes by articulating implication of the approach in any contemporary assessment system.

Keywords: Evaluation, formative assessment practice, learning-centred assessment, validation, validity

Introduction

Classroom-based assessment validation has received critical attention and many conceptualisations have emerged (e.g., Bonner, 2013; Chapelle & Voss, 2013; DeLuca, 2011; Kane, 2016; Kunnan, 2018). While these conceptualisations are helpful in advancing our assessment knowledge, there is a need for a more theoretical conceptualisation of a learning-oriented teacher assessment practice validation. This paper provides a theoretical explication of a classroom-based assessment validation approach as a practical alternative to the traditional validation framework to address how teachers might be engaged to provide context-based evidence for evaluating classroom assessment's practicality, trustworthiness, and usefulness. The current validation approaches have failed to provide learning-oriented evidence for many



reasons, such as the absence of learning-oriented interactions between evaluators, users of assessment data, and students. Also, traditional validation approaches are influenced mainly by positivist thinking, viewing learning as a structural activity, and using the deductive method to provide evidence for the claims about learning, seen as an ordered phenomenon (Kaboub, 2008). Another shortcoming of traditional approaches is the need for more scaffolding for teachers to implement this evaluation in their practice, as teacher education may be insufficient in providing such expertise (Leung, 2015; Pea, 2004).

Based on these limitations, we argue that classroom-based assessment validation should focus on the consequential use of assessment and assessment data based on the conceptual framework of assessment *for* learning (AFL). The classroom-based assessment validation would be the most practical approach, especially when assessment is conceptualised within a broader pedagogical model. Thus, evaluating the effectiveness of assessment should focus on how assessment and assessment data are used to help improve learning and teaching activities. A learning-centred validation approach is needed to account for the various assessment processes that improve learning and teaching. Also, this approach would account for the actual classroom context, with students having individual learning characteristics and needs, and coming from diverse cultural and language backgrounds. This student diversity in the classroom contributes to assessment content and constructs becoming more complex to measure, as these constructs are value-laden with social and cultural implications (DeLuca, 2011).

Given the current understanding of classroom assessment, which entails teachers employing a range of assessments from informal contingent formative assessments to the most formal summative assessment (Davison, 2007; Black & Wiliam, 2018), effective assessment systems should be tailored towards improving learning and teaching. This paper proposes a learning-oriented validation framework to evaluating teacher assessment practices by answering the following questions:

1. What are the limitations of the traditional validation approaches for teacher assessment?
2. How do pragmatic and sociocultural theories address the tensions of balancing accuracy and diversity in validating socio-culturally informed classroom assessment?
3. What evaluation framework could be used to support teachers' assessment evaluation practices and embedding this into teacher professional practices?

Theoretical Framework

The Effective Assessment System

An effective, well-balanced classroom-based assessment system requires a learning-centred validation framework and uses all forms of assessments and multiple sources of evidence (qualitative and quantitative) in supporting learning and teaching activities (Chappuis, et al., 2017; Davison & Michell, 2014). Any form of assessment, from contingent in-class formative

assessment (FA) to the most formal summative assessment (SA), including national and international tests, can be used to support learning and teaching activities and for reporting student outcomes for accountability purposes (Davison, 2007).

The strategic placing of classroom assessment at the centre of instruction and pedagogy is widely noted in an educational system due to its impact on learning and teaching (Black & Wiliam, 1998, 2018). However, this impact seems limited, as the current debate highlights teachers' challenges in using formative and summative assessment data for learning and teaching (Mosher & Heritage, 2017; Yan, et al., 2022). Using assessment *for* learning and teaching is the very essence of postulating a learning-oriented assessment influenced by principles of assessment *for* learning (A \mathcal{L}) (Assessment Reform Group, 2002). The A \mathcal{L} principles are increasingly taken up to underpin teacher assessment practices, including using SA to enhance teaching and learning (Black, 1993: 199; Black, et al., 2011, 2003; Chappuis, et al., 2017; Davison, 2019; Davison & Leung, 2009; Popham, 2017). This A \mathcal{L} concept differs from the traditional outlook on assessment, which is the assessment of learning (AoL), where assessment is used to determine student achievement. Therefore, an effective assessment system, based on A \mathcal{L} principles, discourages teacher assessment practices that centre teachers using assessment data in a summative way to meet the accountability requirements (Brookhart, 2020; Mosher & Heritage, 2017; Shepard, 2020; Willis & Klenowski, 2018).

Consequently, research in assessment suggests that teachers can only fully understand and implement effective assessment when they recognise the social, political, and economic contribution and impact of their assessment practices and how these factors shape their assessment practices (Gipps, 1999). Through this sociocultural lens, teachers should see assessment as a tool that can be used to create a learning community (Hayward, 2015; Lave, 1991) of students and teachers, which then shapes teacher classroom practices and student learning behaviour. From this perspective, learning is not just the individual cognitive processes but also involves the social processes and cultural settings that influence learning. Similarly, others have highlighted the importance of social interactions, cultural contexts, and the belief systems of both students and teachers and how these factors shape students' identity (Black, et al., 2006; Cowie, 2005; Keppell & Carless, 2006; Marshall & Drummond, 2006; Munns & Woodward, 2006), and the nature of control in the classroom shapes learning (Black & Wiliam, 2009). Although one of the aims of using assessment in the classroom is to guide students to become self-regulated and independent learners, self-regulation is critically dependent on interactions with their teachers and peers to activate and support their learning. The conceptualisation of teacher assessment *for* learning (A \mathcal{L}) literacy from a sociocultural perspective highlights the pivotal place of teachers' beliefs and roles in activating student learning. Irrespective of the types of assessment being used in the classroom, the dynamics of the sociocultural context of learners should be considered when interpreting and using assessment results to improve instruction and promote learning (Gipps, 1999; Nortvedt, et al., 2020).

Given the complexities of an effective classroom-based assessment system, the current validation approaches must provide the framework that accounts for the practicality, usefulness, and trustworthiness of teacher assessment practices. There is a need to reframe it to ensure that teachers can make relevant inferences based on the consequential validity of assessment, where the main aim of using assessment is to improve learning and teaching activities. The following section highlights the limitations of the current validation approaches, arguing their insufficiency in addressing the issues around classroom assessment.

Inadequacy of Traditional Validation Approaches

The current validation approaches are measurement-oriented approaches, which are carried out with psychometric principles by focusing on the consistent measurement of student achievement (Farnsworth, 2013; Kane, 1992, 2001, 2006, 2013; Marion & Pellegrino, 2007; Messick, 1989; Sireci, 2013). Although it started as a systematic process of determining the validity of a test, validation, however, has been approached in different ways since the formal treatise of validity by Frank Freeman, who discussed the 'technique and validity of test methods' in the annual report of the Psychological Bulletin of 1914 and the 1921 call for a consensus meaning of validity by the National Association of Directors of Educational Research (Newton & Shaw, 2014). In this section, we present a snapshot of the theoretical positions underpinning traditional validation, how these apply present-day classroom assessment evaluation and then highlight some of the limitations in applying them to classroom assessment evaluation. Educational assessment validation, therefore, has been influenced by two broad theories – the standard-based validation approach and the argument-based approach.

Standards-based validation theories

Standards are conventions that are usually associated with a formal document developed by an organisation, used for regulating the activities of professionals and as a means of ensuring good practices in a particular profession. Various standards have been promulgated to direct assessment evaluation. The National Association of Directors of Educational Research led the first formal attempt to set standards in the design, development, and implementation of educational measurement. However, the widely used document is the *Standards for Educational and Psychological Testing* (henceforth Standards), by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), which have gone through different editions (1966, 1974, 1985, 1999, and 2014).

Standards for assessment validation were informed by validity theory, considered as the extent to which a test measures what it was supposed to measure (Newton & Shaw, 2014). This understanding ensures evidence is provided for physical quality and content accuracy to support conclusions from an assessment. This approach limited the potential of using assessment to support teaching and improve learning, while the approach to investigating

validity was inadequate due to how it was conceptualised (Xi & Sawaki, 2017). Evidence to validate assessment was not informed by test users, particularly teachers and learners, but only by undertaking correlational and test item analyses (Xi & Sawaki, 2017).

Secondly, the trinitarian theory of validity directs all measurement-oriented validation by providing evidence for the content validity, criterion validity, and construct validity as determined by the combined effort of the AERA, APA, and NCME. The trinitarian framework involves investigating the 'validity (face validity, validity by content, validation of the conditions required to answer test items, and empirical validation, namely, concurrent and criterion-based validation) and reliability' (Kunnan, 2018: 37). Although the articulation of the trinitarian framework was commendable, yet this approach heavily relied on providing the psychometric properties of the test by using some advanced statistical methods, which were neither within the professional competency of classroom teachers nor ingrained into the teacher education curriculum (Leung, 2015; Xi & Sawaki, 2017).

Another validity theory is the unified construct approach that addresses the multifaceted, trinitarian approach and shifts the focus from test, to test score interpretation and uses, as identified in the subsequent Standards (1985, 1999 and 2014). Thus, validity is seen as 'the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores' ; at the same time, test validation was taken to be 'the process of accumulating evidence to support such inferences' (AERA, APA, & NCME, 1985: 9). Kunnan (2018: 37) describes it as an 'expanded conceptual framework of validity that included facets of validity of test-score interpretation in terms of values and social consequences of tests and testing' . However, this approach was not without criticism. For instance, the 1985 Standards lacked explicit discussion of 'modern psychometric issues, such as structural equation modelling, function and model fit statistical comparisons of reliability, item response theory, generalizability theory, or computerised testing applications' (Boyle, 1987: 236). There was also the concept of test fairness that could not be fully addressed in situations such as using assessment for teaching and learning and test truthfulness in serving diverse examinees, thereby questioning the fairness of tests in different social situations (Kunnan, 2018).

Moreover, it raises the issue of how a construct should be defined for various learners with different learning needs and ability levels, particularly in a diverse classroom, where examinees come from diverse socio-cultural backgrounds and assessment is tailored to individual student needs. Therefore, there is a need to contextualise the construct and the assessment language, as tests for a particular group may not be suitable in a different context. Thus, specifying the interpretations of the use of a test is somewhat a concern in today' s classroom assessment, as many issues arise around the social and political implications of test use as well as various understandings of language ability informed by the contexts of its uses (Bachman, 2013). Given the need to contextualise proficiency, a validation study of assessment use should be based on the purpose of the assessment. In supporting assessment score interpretation and use, the consequential validity would require triangulation of data sources, as rightly argued by Chan (2014: 12):

... no singular source of evidence sufficient to support a validity claim. Construct validity is the central component in validation work, encompasses the following five sources of evidence germane to the validation of the interpretation and use of the score of an instrument. The five sources include (1) evidence based on test content, (2) evidence based on response processes, (3) evidence based on internal structure, (4) evidence based on relations to other variables, and (5) consequences.

Lastly, these interpretations of test scores are informed by using statistical methods and psychometric evidence of the test items. Due to the very sophisticated process of the Standards, teachers cannot implement the Standards by themselves but rely heavily on external evidence provided by assessment developers. Thus, the Standards need to be clearer on how to evaluate teacher-based assessments. Given this scientific process required in the Standards, some experts have prescribed an argument-based approach, which, however, still needs to fully provide teachers with a fuller understanding of a classroom-based assessment evaluation. The subsequent discussion highlights the works of authors who have prescribed argument-based validity and applied the framework in assessment evaluation.

The argument-based approach

The argument-based approach to evaluating assessment has been widely used for many years since Kane's (1992, 2001, 2006, 2013, 2016) explication of the framework. The argument-based framework is preferred among assessment evaluation researchers (Chapelle & Voss, 2013; Xi & Sawaki, 2017). This approach was first cited by Cronbach (1988), who considers validation an evaluative argument based on five perspectives: functional, political, operationist, economic, and explanatory, and validators as debaters. These views about assessment validation are critical to the present-day evaluation study. For example, a functional perspective to assessment evaluation is instructive – the view about test scores being solely absolute in decision-making should be debunked and a context-based view of test score interpretation is critical in the present-day educational system. Any decision based on the test score only can become dysfunctional in many contexts and, thus, bring about negative consequences (Cronbach, 1988; Messick, 1980). Therefore, validation is considered an evaluative argument that takes into consideration the potentially diverse assessment stakeholders. It 'must link concepts, evidence, social consequences, and values ... an affirmative argument should make clear and, to the extent possible, persuasive the construction of reality and the value weightings implicit in a test and its application' (Cronbach, 1988: 4-5). While the social dimension of assessment is critical (McNamara & Roeber, 2006); however, a limitation to this approach is how teachers are articulated in the framework to take on the responsibility of accumulating evidence in the validation process.

The argument-based evaluation requires an integrative, evaluative judgment of the degree to which evidence and theoretical rationales support the adequacy and appropriateness

of inferences and actions based on test scores. As such, validity is an inductive summary of both the existing evidence for and the potential consequences of test interpretation and use. Hence, what is to be validated is not the test as such, but the inferences derived from the scores - inferences about score meaning or interpretation and about the implications for action that the interpretation entails (Messick, 1987: 1; Messick, 1989: 13).

Furthermore, assessment validation entails the use of inferences, which is the hypotheses made about test score interpretation and uses. The process of answering these questions (inferences), i.e., providing evidence for these inferences, is validation. Messick (1987: 6) describes the evidence as 'both data, or facts, and the rationale or arguments that cement those facts into a justification of test-score inferences'. Expectedly, the sources of this evidence are limitless and are determined by the type of inference made about test score interpretation and uses. These sources could be identified when evaluating the content of the domain of reference or analysing the relationship between individual test responses and assessment tasks. Still, sources of evidence can be informed by the internal structure of test responses and the external background factors of the examinees. Other sources come from evaluating test processes and structure through repeated testing, establishing the relevance of the test based on instructional and remedial implications as well as motivation. Lastly, evaluation could mean tracing the social consequences of test scores interpretation and uses by reporting the intended and unintended outcomes of the test (Messick, 1987, 1989).

Also, the argument-based framework is based on Toulmin's argumentation model (Toulmin, 1958, 2003; Toulmin, et al., 1979, 1984). In this approach, an argument structure is formulated to guide validation studies of educational assessments (Chapelle, et al., 2008; Fulcher & Davidson, 2009; Mislavy, 2003). Elsewhere, this approach is referred to as the interpretive argument and the interpretation and use argument (IUA) framework (Kane, 1992, 2001, 2006, 2013). According to this framework, assessment evaluation integrates score interpretations and uses because claims about assessment scores are based on the interpretation of such scores and uses, which require 'decisions about these units of analysis' (Kane, 2013: 2). Kane's framework uses Toulmin's argument layout that defines the logical, functional procession in the argumentation process, containing six parts: claim, grounds, warrant, qualifier, rebuttal, and backing. Claims from assessment score interpretation and uses are based on grounds (i.e., assessment data) that require warrants. The warrants in the argument-based framework are made stronger by backings. The process requires scrutiny of the grounds (data) upon which the claim is made by providing qualifiers to accept the claim or using rebuttals in rejecting the claim. In Kane's view, the types of evidence in the interpretive argument are observation, generalisation, extrapolation, theory-based implications of decisions made about test scores, and technical issues (Kane, 1992).

Given the relevance of the interpretation and use argument (IUA) framework to evaluate the plausibility of the assessment score, providing a balance between score interpretation and score uses (Kane, 2006, 2013), this approach to classroom assessment has limitations. Bachman and Palmer (2010) objected to Kane's view about evaluating test score interpretation and

uses, as the argument approach was perceived to be too specific, and the claims made exclude emerging usefulness of an assessment in different contexts. A context-specific framework is critical to providing evidence of assessment evaluation (Xi & Davis, 2016). Another shortcoming of the argument-based framework is the accessibility of the framework due to the complex levels and the sophisticated construction of the argument (Bachman & Palmer, 2010). Evaluators, particularly teachers, are yet acquainted with some of the intricacies of implementing an argument-based validation. Therefore, a practical, simplified framework is needed to serve teachers in evaluating their assessment strategies and practices.

To simplify the contextuality, complexity and sophistication of implementing an argument-based validation, another framework was proposed - Assessment use argument (AUA) (Bachman & Palmer, 2010). The AUA is significant in evaluation studies as it helps streamline assessment into justifying the uses of assessment scores and evaluating the potential impacts and consequences of assessment. However, ignoring the interpretation of test scores and focusing assessment evaluation on the uses of test scores can limit the essence of accumulating evidence of the claims made. Both interpretation and uses of test scores are interwoven and potentially challenging to separate in the assessment process (Cronbach, 1988; Kane, 2013).

Lastly, argument-based frameworks are limited, like the Standards; hence, many of their applications were in other settings besides classroom. For instance, the argument-based framework, with formulated inferences, was used to evaluate the Test of English as a Foreign Language (TOEFL) (Chappelle, et al., 2008). Chappelle, et al.' s (2008) inferences (domain description, evaluation, generalisation, explanation, extrapolation, and utilization) can be adapted to suit the claims being made in the classroom context. Hence the methodology is different for a learning-oriented assessment evaluation. The argument-based validation approach strongly focuses on high-stakes standardised assessment and fails to consider the dynamics of classroom assessments.

Given the limitations of the various approaches to evaluating the validity of teacher assessment practices, we explore the potential use of pragmatic and sociocultural theories to guide a learning-centred validation of teacher assessment practices.

The Potential Use of Sociocultural and Pragmatic Theories

This part explains two different theories – the sociocultural theory of learning and the pragmatic approach and what they offer in setting up a learning-oriented argument-based framework suitable for evaluating teacher assessment. These two theories shape our understanding of what should be an essential consideration during validation activities. The sociocultural aspect of assessment highlights the concerns about the limitations of traditional assessment validation approaches, which are not suitable for use by teachers in a multicultural classroom. Then, we explain that the pragmatic principles offer an opportunity to address the limitations of the traditional validation approaches.

Sociocultural theory

The concept of Sociocultural theory (SCT) (Penuel & Wertsch, 1995; Wertsch, 1985) was implied from the work of Vygotsky whose idea of learning and education is explained from a cultural-historical perspective (Fleer, 2015). SCT concerns the interplay of human cognition and social processes or cultural artifacts (Lantolf, et al., 2015). Social factors contribute significantly to learning activities. Learning is, therefore, dependent on the prior existence of more complex cognitive structures, but these more complex cognitive structures are situated in the culture, not in the child. The child acquires them through interaction with adults, who help the child do things that it could not do alone. Through such shared activities, the child internalises the cognitive structures necessary to carry on independently (Bereiter, 1985: 206).

According to this theory, human development is evaluated 'through participation in cultural, linguistic and historically formed settings, such as family life, peer group interaction, and in institutional contexts like schooling, organised social activities, and workplaces ...' (Lantolf, et al., 2015: 1).

The application of SCT to educational activities, especially in the practice of educational assessment, has been argued for in many studies (Brookhart & Helena, 2003; Gipps, 1999; Mislavy, 2008; Moss, 2008; Moss, et al., 2008; Shepard, 2000b; Shepard, et al., 2018b). Therefore, the tenets have informed the design, development, and validation of classroom-based assessment and language assessments. Further, the consciousness of these tenets is evident in the past and in the most recent debates about using assessment to promote learning and teaching and student engagement (Black & William, 1998, 2004, 2018; Shepard, et al., 2018a; Stiggins, 2005). Moss (2008) and Moss, et al. (2008) studied the implications of the SCT on classroom assessment.

SCT can help develop a context-based validation approach to address the validity of using assessment to promote teaching and learning by providing context-based claims on assessment use (Bachman, 2013; Bachman & Palmer, 2010). The sociocultural theory advocates for meaningful and democratic interactions to promote learning. Hence, an important purpose of classroom assessment is using the assessment to promote healthy teacher-student and peer-to-peer interactions. Also, SCT helps to develop a framework to scaffold teacher understanding of assessment evaluation and simplify the complex levels involved in the traditional argument-based framework (Pea, 2004; Xi & Sawaki, 2017). Thus, the learning-centred framework for evaluating classroom-based assessment promotes teacher learning through their engagement in the evaluation process. It is believed that learning is not just a structural activity but a result of many cultural interplays and awareness of self in constructing learning. As an essential principle of SCT, the framework can serve as a scaffold to enhance teacher understanding and evaluation of teacher assessment practices. Involving a teacher in the evaluation process is deemed more practicable than leaving the process entirely in the hands of assessment specialists. The diversity in classrooms requires a conscious harnessing of the social and cultural resources within the classroom setting to accumulate context-based

evidence of an assessment in enhancing teaching and learning. The following section briefly discusses how pragmatism, as a philosophical stance, contributes positively to proposing a framework for classroom assessment validation.

Pragmatism in assessment validation

This section evaluates pragmatism and borrows from past theoretical discussions from pragmatists, to justify suggesting a learning-oriented validation approach. Using pragmatism in assessment validation attempts to answer the most central question about assessment - “How might we provide evidence for evaluating the practicality, trustworthiness and usefulness of classroom assessment or even externally provided assessment?” . What, then, does pragmatism bring to a teacher-based validation framework? The immediate answer would be whatever is feasible for teachers and applicable to justify assessment *for* learning theory. However, further explanations will be provided to support the contribution of pragmatism.

In applying a pragmatic approach, we need to consider the meaning of pragmatism as a branch of philosophy and past theorisation about pragmatism in validation research. Pragmatism finds a balance between ‘ an empirical philosophy that is not religious enough, and a religious philosophy that is not empirical enough’ (James, 1907: 15). In defining pragmatism, James (1907) cited Charles Peirce, also a pragmatist philosopher, who first introduced pragmatism into philosophy. The following extract of Charles Peirce’ s response to pragmatism summarises the very principle of pragmatism:

to develop a thought’ s meaning, we need only determine what conduct it is fitted to produce: that conduct is for us its sole significance. And the tangible fact at the root of all our thought-distinctions, however subtle, is that there is no one of them so fine as to consist in anything but a possible difference of practice. To attain perfect clearness in our thoughts of an object, then, we need only consider what conceivable effects of a practical kind the object may involve – what sensations we are to expect from it, and what reactions we must prepare. Our conception of these effects, whether immediate or remote, is then for us the whole of our conception of the object, so far as that conception has positive significance at all. (Cited in James, 1907: 46-47)

From this explanation, the understanding of truth is how transferable the truth is in real-life application. Thus, ‘pragmatism seeks to account for how people make sense of the world through action in the world’ (Stone & Zumbo, 2016: 557). Through the proposed framework, teachers could help to establish the meaningfulness of assessment in enhancing their teaching and promoting learning, using their cultural and social knowledge of the classroom as a resource, contributing to an appropriate approach to classroom assessment evaluation.

Within this philosophical understanding, we provide a “possible difference of practice” in the evaluation of a learning-oriented assessment. The purpose is to provide an accessible framework with which teachers can easily connect. The technicality of most past evaluation

practices is far-fetched for the classroom teacher leaving the evaluation the sole responsibility of assessment specialists. Unlike many measurement-oriented validation studies, this pragmatist, learning-oriented approach can promote stakeholders' interaction with assessments. We suggest this approach, taking inspiration from Stone and Zumbo (2016), who reaffirmed the potentially global usefulness of pragmatist thinking in accumulating evidence for test use and as a possible fit for addressing local issues. Thus, learning-oriented assessment evaluation would require a specific approach informed by pragmatic principles, as Stone and Zumbo (2016) highlighted. Their four principles of a pragmatic approach can be summarised: Acceptable actions should proceed with ease. Any impediments to our actions require us to chart new courses of action to make a distinction. Like empiricism, pragmatism appeals to senses; what is considered valuable is the meaning we make concerning the particularity of our actions, contexts, and situations; what works well for the context is the primary concern of the pragmatic approach (Stone & Zumbo, 2016).

Given that 'pragmatism is concerned with what works in practice' ; and theories and concepts 'are instruments that we use to guide our actions' (Stone & Zumbo, 2016: 559-560), we argue that a practical approach to classroom assessment evaluation should involve teachers. Teacher engagement is critical as they possess a socio-cultural understanding of the classroom and can constantly review the "possible difference of practice" through reflections and democratic interactions between the teachers and their students regarding teaching and learning. We believe this teacher reflection on practice can provide a pragmatic way to address the tension of balancing assessment usefulness trustworthiness, and diversity in a modern classroom.

With the potential of sociocultural and pragmatic theories to address the tensions of balancing accuracy and diversity in validating teacher assessment practices and to account for the co-construction nature of learning among the stakeholders, we propose a context-based validation approach for evaluating teacher assessment practices that is practical, trustworthy, and useful and as an alternative to the standards and argument-based approaches.

The Proposed Learning-centred Assessment Evaluation Framework

This section describes a learning-centred assessment evaluation framework underpinned by sociocultural learning theory. It seeks to provide an alternative to traditional approaches, expanding some of the tenets of educational assessment evaluation in the Standards as well as the argument-based framework. First, we highlight important principles guiding assessment evaluation and discuss appropriate hypotheses (inferences) related to classroom assessment practices. Then, we explain how this approach to evaluation can enhance teacher assessment learning and improve assessment *for* learning practices. Finally, we draw on past explanations about classroom assessment practices and validity theories to construct a teacher-based validity argument by highlighting claims from score interpretations and uses relevant to classroom assessments. This section would be a step towards using an adaptable framework for evaluating teacher-based assessment practices because, presently, 'despite an extensive

tradition and literature on validation of standardised educational and psychological tests, appropriate methods for validation of assessments used by teachers in classrooms are not well defined' (Bonner, 2013: 87).

Key principles and inferences for evaluating classroom assessment

There are guiding principles that guide the psychometric approach to validity. Some of these principles apply to classroom assessment evaluation. This section briefly provides practical tenets of applying validation theory to classroom assessments. We discuss four useful principles for a learning-centred validation framework.

First, the foundational principle for classroom assessment design is that it is built for learning purposes. One criticism of the traditional assessment validation approach is that it focuses on measurement, i.e., measuring student ability, progress, and skill. In contrast, we propose a learning-centred assessment evaluation practice using assessment to enhance learning and teaching activities. Decisions made from score uses are purely for learning purposes. Further, in this framework, plausible inferences for a test score interpretation and use in a classroom context are domain definitions, evaluation, explanation, utilisation, and ramification inferences. It is assumed that both generalisations and extrapolation inferences are unlikely to address claims relating to the uses of assessment to improve learning and teaching. Instead, both assumptions focus on the comparison of test scores interpretation and use with criteria external to classroom contexts. The purpose of using assessment to predict future performance is antithetical to assessment *for* learning principles (Shepard, et al., 2018b). Learning is not structural but is expressed as developmental progress, referred to as a learning progression (Briggs & Furtak, 2020; Lehrer & Schauble, 2015). Teacher perception of learning as a progression is critical to evaluating assessment claims.

Second, the teacher-made inference is made deductively and based on the teacher's conceptual framework (Black, et al., 2010; Bonner, 2013). However, conceptualisation of classroom assessment primarily relates to claims which address where a student would need to reach curriculum expectations; where learners are already; how best to help students reach learning goals (Black & Wiliam, 2018). To address these claims, relevant inferences teacher could make about test score interpretations and uses are domain definitions, evaluation, explanation, utilisation, and ramification. In addition to teachers developing inferences before using an assessment approach, this inference could also be set for teachers by external institutions, especially for externally developed assessments for formative purposes (Chapelle, et al., 2015; Koizumi, 2015). It is important to note that externally produced assessment should employ a bottom-top approach in the design and development of assessment, including setting purposes of assessment. What this means to a learning-centred assessment validation is that the setting up of the interpretive argument, which usually takes place in the development phase of assessment, should integrate teacher-own purposes, and intended use of the assessment. The more teachers are involved in the test design and development, the more substantial the evidence a teacher provides in the evaluation of the assessment (Bonner, 2013).

The third principle is that the purposes of assessment evolve. For an emerging interpretation, purpose and use, new validation should be carried out to address its context. Hence, evaluation can be carried out formatively (Kane, 2006) due to the fluid nature of teacher assessment practices (Bonner, 2013). Also, all possible sources of evidence (quantitative and qualitative) should be explored by all assessment stakeholders before attempting to use assessment data (Bachman & Palmer, 2010). There should be a triangulation of sources to either accept or reject assessment claims. A teacher works towards accumulating evidence to approve or disapprove the claims. In an instance of disapproval, evidence of rebuttals to the claims should be provided to develop a holistic argument for assessment score interpretation and use (Kane, 2006, 2013).

Lastly, the validity argument must be time-sensitive in responding to assessment tasks. A teacher can seek immediate feedback from the student about the cognitive processes undertaken by the students. Some evidence is collected during test administration, while some are best collected after test administration. Therefore, teacher involvement in assessment evaluation is a continuous and recursive approach to assessment evaluation. Evaluation is never a one-time effort but a circular process from planning the assessment process to developing and designing the assessment. A teacher-based evaluation approach should be an embedded process built into daily activities of teacher assessment practices. The discussion section of this paper provides how the framework can be used to develop teacher assessment evaluation literacy. Next, we present the framework for evaluating classroom assessment.

The learning-centred evaluation framework

This section presents the learning-centred framework (Figure 1) for evaluating teacher assessment practices. As argued above, the framework's purpose is to provide a practical approach to assessment validation, which can be entrenched in teacher assessment praxis. The framework relies on pragmatic philosophy and supports a socio-cultural, learning-oriented evaluation approach as argued for in several classroom-based assessment literatures (Brookhart & Helena, 2003; Gipps, 1999; Moss, 2003, 2008, 2016; Moss, et al., 2008; Shepard, 2000b, 2009).

The learning-centred framework is drawn from the extant theories on validation framework, such as the one presented by Chapelle, et al. (2015). However, we provide a more adaptable model to evaluating teacher assessment practices. As argued, we use socio-cultural and pragmatism theories to develop this framework, evident in our outline of the inferences, warrants and claims considering the context-specific teacher assessment practices.

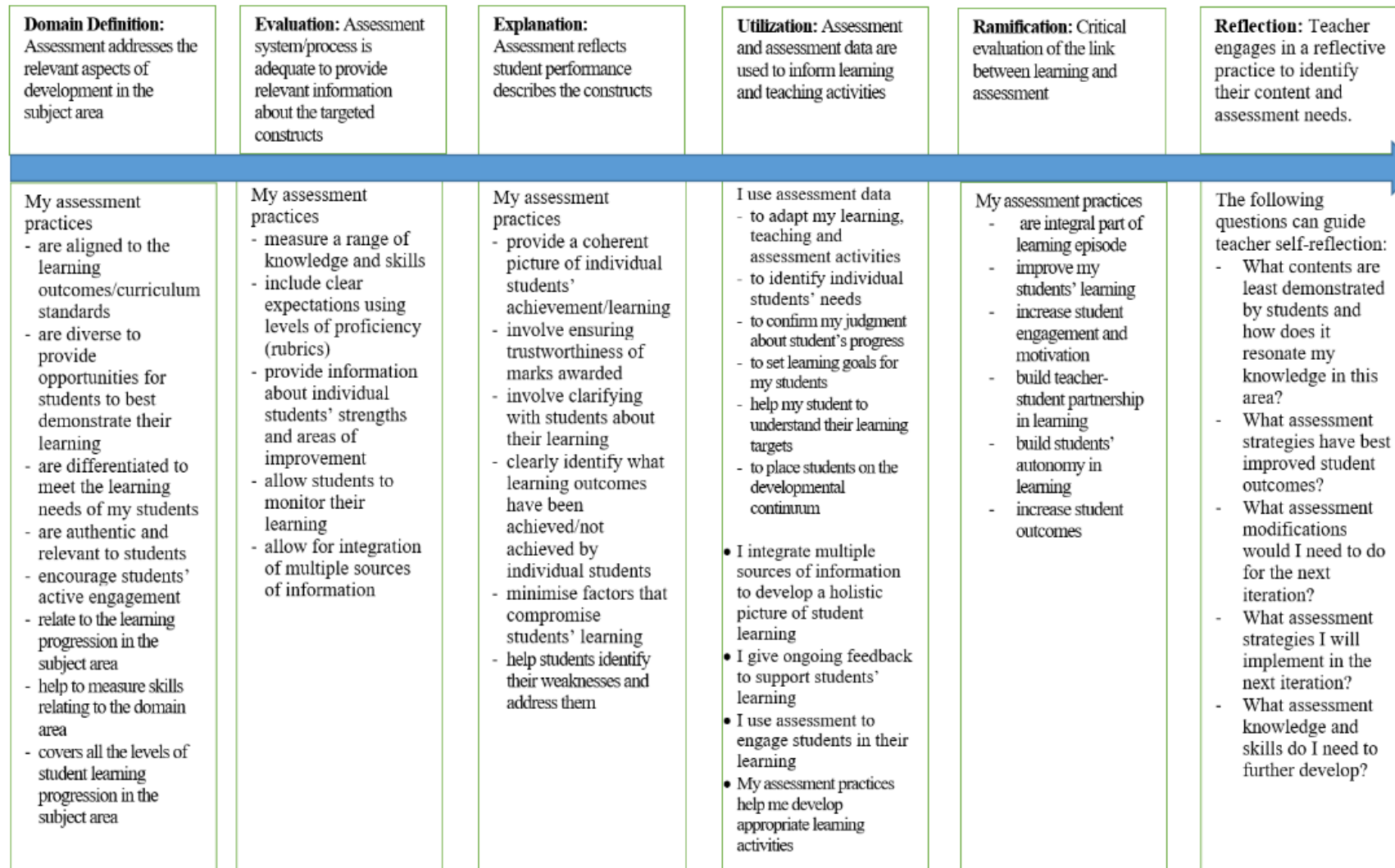


Figure 1. Inferences, warrants, and claims in the proposed learning-centred assessment evaluation framework.

To undertake a learning-oriented assessment evaluation, we articulate six inferences relevant to plausible claims from assessment score interpretation and use, based on the principles of effective assessment practices. First, we use a domain definition to set important considerations in classroom assessment. Domain definition inference supports teachers in ensuring that assessment tasks align with the relevant aspects of curriculum expectations and are suitable for the context of classroom assessment. We highlight some of the claims in the proposed framework, requiring evidence to evaluate teacher assessment practices based on domain definition inference.

The second inference is the evaluation inference, which helps to collect evidence regarding how learners are evaluated, the scoring process, and linking the evaluation stage to the appropriate curriculum content and learners' characteristics.

Explanation seeks further confirmation about the quality of teacher assessment practice, prompting teachers to gather evidence on the strategies learners used in undertaking assessment tasks. The explanation inference also helps to connect student performance with the construct being assessed.

Utilisation and ramification inferences are critical to evaluating classroom assessment practices (Chapelle, et al., 2008, 2015), though there is only a thin line between the two inferences. While utilisation inference deals with decisions made because of the assessment result, the ramification inference considers those claims about improved learning when an assessment result is used, especially by a student. Teacher' s reflective practice provides evidence of assessment use. In other words, the reflection inference ensures that assessment practice promotes self-reflection for teachers.

Following each inference, a warrant leads teachers to the claims to support assessment practice. The claims are numbered under each warrant and are not in any way exhaustive. The claims presented are speculative of what we consider could apply to the intended purposes of and theoretical explanations of classroom assessment practices in the literature (e.g., Black & Wiliam, 2018; Chapelle, et al., 2015, 2008). Lastly, it is assumed that the evidence required for each claim is within reach of classroom teachers, however, we acknowledge that there are challenges teacher may encounter in using the learning-oriented framework. For instance, teacher cultural orientation about assessment may hinder student autonomy in learning, discouraging student-led assessment approaches. Another issue is the culture of accountability in many jurisdictions, which may hinder teacher to imbibe the principles of learning-oriented assessment evaluation. There are several factors that could impact teacher involvement in this framework, which we do not intend to address in this exploration, but we believe, when applied, the framework contributes positively to the way classroom-based assessment evaluation is implemented. In the next section, we discuss the contributions of this framework to the field of classroom assessment practice and research.

Discussion

The section discusses the contributions of the framework to teacher assessment practices and

professional development. There are six contributions to the field of classroom assessment and educational evaluation that our paper seeks.

First, we extended the current conceptualisation of validation approach that accounts for the socio-cultural context of assessment. A learning-oriented validation framework advocates giving attention to the context of assessment use in evaluating teacher assessment practices. Many of the evaluation studies were mostly carried out by external evaluators. However, the teacher as the evaluator can understand the primary purpose of using assessments in their context and justify the context of student learning, i.e., using student and the classroom context 'for making inferences about student learning' (Black & Wiliam, 2018: 2), as the context of assessment differs from one classroom to another. In the same way, justifiable use of externally developed assessments can be established based on teacher evaluation of such assessment use in the classroom.

Second, through the framework, we have provided a way by which validation of assessment practices becomes an integral component of teacher assessment practices. Given that the teacher education process is limited in equipping classroom teacher' s understanding of what assessment is, not to mention evaluating the process and has resulted in a narrow-minded classroom assessment practice (Hamp-Lyons, 2007), the framework can be a reflective tool to improve assessment practices. Hence, we argue the need to imbibe in teachers a process of evaluating their assessment practices. This study helps to fill this gap in teacher assessment education and argues for integrating assessment evaluation into teacher assessment practice, viewed as a means to an end and not as an end in the teaching and learning process. Further, the framework can best be implemented as part of a teacher' s professional responsibility for ensuring an effective assessment practice. With the framework, a teacher can provide evidence of intended and unintended impacts of an assessment process. This idea of embeddedness follows the scholarly argument of best assessment practice, which is integrating assessment into teachers' instructional activities (Black & Wiliam, 1998, 2018; Shohamy, 1998; Wiliam & Leahy, 2015).

Thirdly, the framework advances teacher assessment *for* learning (A \mathcal{L}) practices (Assessment Reform Group, 2002) and ensures teachers use assessment to promote student learning. As this paper advocates, the tenets of A \mathcal{L} are entrenched in the learning-centred framework that can help to promote teacher theory of assessment evaluation in undertaking a learning-focused assessment practice. Thus, the framework guides against the negative washback of AoL practices, consistently reported in the literature (Alderson & Wall, 1993; Kunnan, 2010, 2018; McNamara & Ryan, 2011). We have used this framework to argue that the focus of assessment validation is student learning rather than the validity of the measurement (Bonner, 2013; Shepard, 2000a). Claims that support improved learning should be supported appropriately by teacher evaluation of the assessment process. Assessment should support the teacher in the important facets of student learning. Essentially, teachers use assessment to answer crucial questions of where a learner needs to get to, where a learner is currently, and how best to get a learner to the next level. These three questions are critical aspects of A \mathcal{L}

strategies (Black & Wiliam, 2018). Thus, the teacher assessment evaluation framework developed in this paper is informed by the traditional measurement theory (argument model) but heavily influenced by the non-measurement (learning) theory. The two theories, already discussed in this paper, inform the learning-centred validation framework (in Figure 1), highlighting the inferences that inform the possible claims that could be made on using an assessment in promoting student learning.

The framework supports previous conceptualisations of teacher-led evaluation practice. As Bonner (2013: 103) argues that ‘teacher self-inquiry may be more effective than researcher-driven inquiry for improving the validity of teacher practice’. The more a teacher uses the framework, the more they develop expertise, such as communicating effectively, motivating students, and enhancing other assessment strategies (Zhao, et al., 2018). Similarly, the fundamental logic behind a learning-centred assessment validation concerns the teacher’s ability to provide a defensible assessment practice. Literature also suggests that teacher learning can be improved when a teacher ‘engage(s) in continuous (ongoing) classroom assessment integrated with teaching, in a potentially formative role’ (Ratnam & Tharu, 2018: 121). Teacher-based evaluation can contribute positively to enhancing teacher assessment knowledge, beliefs, and orientation, thus promoting teacher ‘understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions’ (Popham, 2017: 25). Teacher participation in the assessment evaluation process is another way of improving assessment expertise (Davison, 2019; Michell & Davison, 2019). As Lam (2018: 115) noted, teacher learning is a ‘complex process, which is beyond the provision of initial teacher education, short-lived professional development ...’. Teacher-led assessment evaluation promotes teacher understanding of their evaluative function of providing ‘the kinds of data called for in the standard to support the valid interpretations of the test results for their intended purposes’ (AERA, et al., 2014: 4). As a teacher gets involved in the study of assessment evaluation, the search for evidence to justify the consistency and accuracy of her/his judgment of the learner’s progress can be actualised. Moreso, what an assessment says about such a learner, is a way of confirming or disconfirming teacher judgment, which in turn gives more confidence to a teacher. In Australia, teachers regularly judge students’ progress towards state and national curricula (Willis & Klenowski, 2018). Therefore, prompts from the evaluation framework can ‘sustain extensive teacher learning about assessment, but also curriculum, instruction, their students, and about themselves’ (Hill, et al., 2018: 195). Hopefully, teacher learning can be improved through a teacher-led assessment evaluation framework. The framework proposed in this paper can help to usher a teacher into a conscious consideration of the potential areas of importance in evaluating assessment practice and a balanced assessment system (Black, 1993; Black, et al., 2011; Chappuis, et al., 2017).

Inclusive assessment and democratic participation of assessment stakeholders can be enhanced using the proposed framework. Teacher assessment practice involves determining a student's current level, as progress towards curriculum expectations (Black & Wiliam, 2018), in a democratic atmosphere. Also, teacher practice is concerned with using assessment to identify

specific strengths and weaknesses of learners and using assessment outcomes to inform instructional decisions (Alderson, et al., 2015; Zhao, et al., 2018). Using the framework, a teacher gets to articulate assessment responsibility more clearly. Also, searching for evidence through student feedback and peer feedback can enhance positive interactions in the classroom (Asadi, et al., 2017; Black & Wiliam, 2018; Hattie & Timperley, 2007). This active participation of teacher and student in assessment evaluation promotes dialogic interaction between a teacher and students (Dixon & Hawe, 2018; Sadler, 2010). For instance, a teacher's investigation into the test takers' cognitive functioning can be an important source of evidence to back a claim based on explanation inference. Some of the claims in the framework would require the teacher to ask questions from the student and take questions from learners regarding the assessment process. More learning-based evidence is provided through teacher-student relationship and using assessment strategies, such as simple questioning, student reflective writing and other advanced techniques to explain test-takers' cognitive processes while evaluating the usefulness and practicality of assessment process.


The framework can be a valuable way to evaluate an externally designed summative-based formative assessment, with teachers responding to the claims in the framework. Teacher responses to the claims can then be used as sources of evidence for the usefulness, trustworthiness, and practicality of an assessment in enhancing teacher assessment practices. The claims raised in the framework are inexhaustible, just a guide. They were informed by theories in the field of classroom assessment and the result of an exploratory qualitative phase during a doctoral study and recently used to evaluate teacher assessment practice (Ijiwade & Alonzo, 2023). Therefore, the evidence a teacher can provide is wider than the claims addressed in the teacher-based evaluation argument. For instance, a teacher can investigate students' cognitive functioning immediately after the assessment to justify that students' responses to the assessment are not because of a lucky guess. The use of simple questioning, student-reflective writing, and other advanced techniques can be deployed by a teacher to explain the cognitive processes of test-takers (Bonner, 2013). Hopefully, the sources of evidence for the claims would be within reach of every teacher. Evaluating assessments based on this framework would increase teacher confidence about her/his assessment practices and the authenticity of the assessment approach to the classroom context.


Conclusions

Educational assessment validation has evolved over the years. While various approaches have been developed, an argument-based framework for evaluating educational assessments is still popular despite its limitations in addressing many context-based issues. Moreso, the practice and the purpose of assessment has evolved, making the argument-based framework misaligned to classroom assessment evaluation. Hence, a pragmatic philosophy and sociocultural theory must provide an adapted argument framework for implementing a learning-oriented assessment evaluation. Although the learning-centred validation approach proposed in this paper is still in its developmental stage, this concept maybe useful to guide

teachers' assessment practices, mainly using the framework to evaluate their assessment *for* learning practice designed to support teaching and promote student learning. Lastly, the framework could be helpful for test developers and classroom assessment researchers, providing important considerations in developing and evaluating the usefulness, trustworthiness, and practicality of assessments promoting learning and teaching.

Author biographies

Oluwaseun Ijiwade is an alumnus of the University of New South Wales Sydney. He is an Assessment Systems Officer at the Australasian College for Emergency Medicine. He researches K-12 and higher education assessment-related issues and applies measurement and non-measurement theories to deliver national and international assessment programs. 

Dennis Alonzo is a senior lecturer at the University of New South Wales Sydney. He researches the intersections of curriculum, assessment, equity, evaluation of educational programmes, and teacher education and development. He has worked with educational systems and schools nationally and internationally to lead their assessment reforms. 

Reference

- Alderson, J.C., Brunfaut, T. & Harding, L. 2015. Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36(2): 236-260.
- Alderson, J.C. & Wall, D. 1993. Does washback exist? *Applied Linguistics*, 14(2): 115-129.
- American Educational Research Association, American Psychological Association, & National Center on Measurement in Education. (1999). Standards for educational and psychological testing. American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Center on Measurement in Education [AERA, APA, NCME]. (2014). The standards for educational and psychological testing. 4th ed. American Educational Research Association. Available at: <http://www.aera.net/Publications/eBooks> (Accessed: 24 July 2017).
- Asadi, M., Azizinezhad, M. & Ehsani fard, E. 2017. Formative assessment and feedback as predictors of students' engagement. *Journal of Research in Applied Linguistics*, 8(Proceedings of the Fourth International Conference on Language, Discourse and Pragmatics, 2017): 291-298.
- Assessment Reform Group. 2002. Assessment for learning: Research-based principles to guide classroom practice. Cambridge, UK: University of Cambridge School of Education.
- Bachman, L.F. 2013. Ongoing challenges in language assessment. In Kunnan, A.J. (Ed.), *The Companion to Language Assessment*. John Wiley & Sons, Inc., 1568-1604.
- Bachman, L. & Palmer, A. 2010. *Language Assessment in Practice*. Oxford: Oxford University Press.

- Bereiter, C. 1985. Toward a solution of the learning paradox. *Review of Educational Research*, 55(2): 201-226.
- Black, P. 1993. Formative and summative assessment by teachers. *Studies in Science Education*, 21(1): 49-97.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. 2010. Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2): 215-232.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. 2011. Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4): 451-469.
- Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. 2003. *Assessment for Learning: Putting it into Practice*. Reprinted. Open University Press.
- Black, P., McCormick, R., James, M. & Pedder, D. 2006. Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education*, 21(2): 119-132.
- Black, P. & Wiliam, D. 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1): 7-74.
- Black, P. & Wiliam, D. 2004. The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education*, 103(2): 20-50.
- Black, P. & Wiliam, D. 2009. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1): 5-31.
- Black, P. & Wiliam, D. 2018. Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6): 551-575.
- Bonner, S.M. 2013. Validity in classroom assessment: Purposes, properties, and principles. In McMillan, J.H. (Ed.), *SAGE Handbook of Research on Classroom Assessment*. London: SAGE Publications, Inc., 87-106.
- Boyle, G.J. 1987. Review of the (1985) Standards for educational and psychological testing: AERA, APA and NCME. *Australian Journal of Psychology*, 39(1989): 235-237.
- Briggs, D.C. & Furtak, E.M. 2020. Learning progressions and embedded assessment. In Brookhart, S.M. & McMillan, J.H. (Eds.), *Classroom Assessment and Educational Measurement*. London: Routledge, 146-169.
- Brookhart, S.M. 2020. Feedback and measurement. In Brookhart, S.M. & McMillan, J.H. (Eds.), *Classroom Assessment and Educational Measurement*. London: Routledge, 53-78.
- Brookhart, S.M. & Helena, M. 2003. Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4): 5-12.
- Chan, E.K.H. 2014. Standards and guidelines for validation practices: Development and evaluation of measurement instruments. In Zumbo, B.D. & Chan, E.K.H. (Eds.), *Validity and Validation in Social, Behavioral, and Health Sciences*. Volume 54. London: Springer International Publishing, 9-24.
- Chapelle, C.A., Cotos, E. & Lee, J. 2015. Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3): 385-405.

- Chapelle, C.A., Enright, M.K. & Jamieson, J.M. 2008. Test score interpretation and use. In Chapelle, C.A., Enright, M.K. & Jamieson, J.M/ (Eds.), *Building a Validity Argument for the Test of English as a Foreign Language*. London: Routledge, 1-26.
- Chapelle, C.A. & Voss, E. 2013. Evaluation of language tests through validation research. In Kunnan, A.J. (Ed.), *The Companion to Language Assessment*. Wiley Blackwell, 1081-1097.
- Chappuis, S., Commodore, C. & Stiggins, R.J. 2017. *Balanced Assessment Systems: Leadership, Quality, and the Role of Classroom Assessment*. Thousand Oaks, CA: Corwin.
- Cowie, B. 2005. Pupil commentary on assessment for learning. *The Curriculum Journal*, 16(2): 137-151.
- Cronbach, L.J. 1988. Five perspectives on validity argument. In Wainer, H. & Braun, H.I. (Eds.), *Test Validity*. Lawrence Erlbaum Associates, 3-17.
- Davison, C. 2007. Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly*, 4(1): 37-68.
- Davison, C. 2019. Using assessment to enhance learning in English language education. In Gao, X. (Ed.), *Second Handbook of English Language Teaching*. Volume 1. London: Springer International Publishing, 1-22.
- Davison, C. & Leung, C. 2009. Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3): 393-415.
- Davison, C. & Michell, M. 2014. EAL assessment: What do Australian teachers want? *TESOL in Context*, 24(2): 51.
- DeLuca, C. 2011. Interpretive validity theory: Mapping a methodology for validating educational assessments. *Educational Research*, 53(3): 303-320.
- Dixon, H.R. & Hawe, E. 2018. Developing assessment-capable teachers through engagement in assessment for learning (AFL): A New Zealand study. In Jiang, H. & Hill, M.F. (Eds.), *Teacher Learning with Classroom Assessment: Perspectives from Asia Pacific*. London: Springer, 59-76.
- Farnsworth, T.L. 2013. An investigation into the validity of the TOEFL iBT speaking test for international teaching assistant certification. *Language Assessment Quarterly*, 10(3): 274-291.
- Fleer, M. 2015. Developing an assessment pedagogy: The tensions and struggles in re-theorising assessment from a cultural– historical perspective. *Assessment in Education: Principles, Policy & Practice*, 22(2): 224-246.
- Fulcher, G. & Davidson, F. 2009. Test architecture, test retrofit. *Language Testing*, 26(1): 123–144.
- Gipps, C. 1999. Socio-cultural aspects of assessment. *Review of Research in Education*, 24(1): 355-392.
- Hamp-Lyons, L. 2007. The impact of testing practices on teaching: Ideologies and alternatives. In Cummins, J. & Davison, C. (Eds.), *International Handbook of English Language Teaching*. London: Springer, 487-504.
- Hattie, J.A. & Timperley, H. 2007. The power of feedback. *Review of Educational Research*,

- 77(1): 81-112.
- Hayward, L. 2015. Assessment is learning: The preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, 22(1): 27-43.
- Hill, M.F., Tan, K. & Jiang, H. 2018. Classroom assessment for teacher learning and student learning. In Jiang, H. & Hill, M.F. (Eds.), *Teacher Learning with Classroom Assessment: Perspectives from Asia Pacific*. London: Springer, 193-201.
- James, W. 1907. *Pragmatism: A New Name for Some Old Way of Thinking*. Longman, Green and Co.
- Ijiwade, O. & Alonzo, D. 2023. Teacher perceptions of the use of a computer-adaptive test for formative purposes: Typologies of practices. *International Journal of Instruction*, 16(2): 88-908.
- Kaboub, F. 2008. Positivist paradigm. In Leong, F. (Ed.), *Encyclopedia of Counselling*. London: SAGE Publications, Inc., 785.
- Kane, M.T. 1992. An argument-based approach to validity. *Psychological Bulletin*, 112(3): 527-535.
- Kane, M.T. 2001. Current concerns in validity theory. *Journal of Educational Measurement*, 38(4): 319-342.
- Kane, M.T. 2006. Validation. In Brennan, R.L. (Ed.), *Educational Measurement*. 4th edn. Praeger Publishers, 17-64.
- Kane, M.T. 2013. Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1): 1-73.
- Kane, M.T. 2016. Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2): 198-211.
- Koizumi, R. 2015. Second language vocabulary assessment studies: Validity evidence and future directions. *Vocabulary Learning and Instruction*, 4(1): 36- 46.
- Kunnan, A.J. 2010. Test fairness and Toulmin' s argument structure. *Language Testing*, 27(2): 183-189.
- Kunnan, A.J. 2018. *Evaluating Language Assessments*. London: Taylor & Francis.
- Lam, R. 2018. Teacher learning of portfolio assessment practices: Testimonies of two writing teachers. In Jiang, H. & Hill, M.F. (Eds.), *Teacher Learning with Classroom Assessment: Perspectives from Asia Pacific*. Berlin: Springer, 99-118.
- Lantolf, J.P., Thorne, S.L. & Poehner, M. 2015. Sociocultural theory and second language development. In van Patten, B. & Williams, J. (Eds.), *Theories in Second Language Acquisition*. London: Routledge, 207-226.
- Lave, J. 1991. Situating learning in communities of practice. In Resnick, L.B., Levine, J.M. & Teasley, S.D. (Eds.), *Perspectives on Socially Shared Cognition*. American Psychological Association, 63-82. Available at: <https://www.ecologyofdesigninhumansystems.com/wp-content/uploads/2012/12/Lave-Situating-learning-in-communities-of-practice.pdf> (Accessed: 19 November 2017).
- Lehrer, R. & Schauble, L. 2015. Learning progressions: The whole world is not a stage. *Science*

- Education*, 99(3): 432-437.
- Leung, C. 2015. Book review—Exploring language assessment and testing: Language in action. *ELT Journal*, 69(1): 109-110.
- Marion, S.F. & Pellegrino, J.W. 2007. A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice*, 25(4): 47-57.
- Marshall, B. & Drummond, M.J. 2006. How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education*, 21(2): 133-149.
- McNamara, T. & Roever, C. 2006. Validity and the social dimension of language testing. *Journal of Research in Language Studies*, 56(S2): 9-42.
- McNamara, T. & Ryan, K. 2011. Fairness versus justice in language testing: The place of English literacy in the Australian citizenship test. *Language Assessment Quarterly*, 8(2): 161-178.
- Messick, S. 1980. Test validity and the ethics of assessment. *American Psychologist*, 35(11): 1012-1027.
- Messick, S. 1987. Validity. *ETS Research Report Series*, 1987(2): i-208.
- Messick, S. 1989. Validity. In Linn, R.L. (Ed.), *Educational Measurement*. 3rd edn. London: Macmillan, 13-103.
- Michell, M. & Davison, C. 2019. 'Bringing the teacher back in' : Toward L2 assessment praxis in English as an additional language education. In Poehner, M.E. & Inbar-Lourie, O. (Eds.), *Toward a Reconceptualization of L2 Classroom Assessment: Praxis and Researcher-Teacher Partnership*. London: Springer, 23-41.
- Mislevy, R.J. 2003. Argument substance and argument structure in educational assessment (CSE Technical Report 605 CSE Technical Report 605). Centre for the Study of Evaluation, CRESST, 1-35.
- Mislevy, R.J. 2008. Issues of structure and issues of scale in assessment from a situative/sociocultural perspective. In Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H. & Young, L.J. (Eds.), *Assessment, Equity, and Opportunity to Learn*. Cambridge: Cambridge University Press, 259-294.
- Mosher, F. & Heritage, M. 2017. A hitchhiker' s guide to thinking about literacy, learning progressions, and instruction. Consortium for Policy Research in Education. Available at: <https://doi.org/10.12698/cpre.RR-2017-2> (Accessed: 23 May 2022).
- Moss, P.A. 2003. Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4): 13- 25.
- Moss, P.A. 2008. Sociocultural implications for assessment I: Classroom assessment. In Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H. & Young, L.J. (Eds.), *Assessment, Equity, and Opportunity to Learn*. Cambridge: Cambridge University Press, 222-258.
- Moss, P.A. 2016. Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy & Practice*, 23(2): 236-251.
- Moss, P.A., Girard, B.J. & Greeno, J.G. 2008. Sociocultural implications for assessment II: Professional learning, evaluation, and accountability. In Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H. & Young, L.J. (Eds.), *Assessment, Equity, and Opportunity to Learn*.

- Cambridge: Cambridge University Press, 295-332.
- Munns, G. & Woodward, H. 2006. Student engagement and student self- assessment: The REAL framework. *Assessment in Education: Principles, Policy & Practice*, 13(2): 193-213.
- National Council on Measurement in Education (NCME). 2019. Revision of the standards for educational and psychological testing. Available at: <http://www.ncme.org/resources/library/standards-revision> (Accessed: 27 August 2019).
- Newton, P.E., & Shaw, S.D. 2014. *Validity in Educational and Psychological Assessment*. London: SAGE Publications Ltd.
- Nortvedt, G.A., Wiese, E., Brown, M., Burns, D., McNamara, G., O' Hara, J., Altrichter, H., Fellner, M., Herzog-Punzenberger, B., Nayir, F. & Taneri, P.O. 2020. Aiding culturally responsive assessment in schools in a globalising world. *Educational Assessment, Evaluation and Accountability*, 32(1): 5-27.
- Pea, R.D. 2004. The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3): 423-451.
- Penuel, W.R & Wertsch, J.V. 1995. Vygotsky and identity formation: A sociocultural approach. *Educational Psychologist*, 30(2): 83-92.
- Popham, W.J. 2017. *Classroom Assessment: What Teachers Need to Know*. 8th edn. London: Pearson Education.
- Ratnam, T. & Tharu, J. 2018. Integrating assessment into classroom instruction to create zones of development for teachers and learners: Some perspectives from India. In Jiang, H. & Hill, M.H. (Eds.), *Teacher Learning with Classroom Assessment: Perspectives from Asia Pacific*. Berlin: Springer, 119-140.
- Sadler, D.R. 2010. Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5): 535-550.
- Shepard, L.A. 2000a. The role of assessment in a learning culture. *Educational Researcher*, 29(7): 4-14.
- Shepard, L.A. 2000b. The role of classroom assessment in teaching and learning. In *Center for the Study of Evaluation National Center for Research on Evaluation, Standards, and Student Testing*. 1-85.
- Shepard, L.A. 2009. Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice*, 28(3) 32-37.
- Shepard, L.A. 2020. Should "measurement" have a role in teacher learning about classroom assessment? In Brookhart, S.M. & McMillan, J.H. (Eds.), *Classroom Assessment and Educational Measurement*. London: Routledge, 192-206.
- Shepard, L.A., Penuel, W.R. & Pellegrino, J.W. 2018a. Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1): 21-34.
- Shepard, L.A., Penuel, W.R. & Pellegrino, J.W. 2018b. Classroom assessment principles to support learning and avoid the harms of testing. *Educational Measurement: Issues and*

- Practice*, 37(1): 52-57.
- Shohamy, E. 1998. Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4): 331-345.
- Sireci, S.G. 2013. Agreeing on validity arguments. *Journal of Educational Measurement*, 50(1): 99-104.
- Stiggins, R.J. 2005. From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4): 324-328.
- Stone, J. & Zumbo, B.D. 2016. Validity as a pragmatist project: A global concern with local application. In Aryadoust, S.V. & Fox, J. (Eds.), *Trends in Language Assessment Research and Practice: The View from the Middle East and the Pacific Rim*. Cambridge: Cambridge Scholars Publishing, 555-573.
- Toulmin, S.E. 1958. *The Uses of Argument*. 1st edn. Cambridge: Cambridge University Press.
- Toulmin, S.E. 2003. *The Uses of Argument: Updated Edition*. Cambridge: Cambridge University Press.
- Toulmin, S.E., Rieke, R.D. & Janik, A. 1979. *An Introduction to Reasoning*. 1st edn. London: Macmillan Publishing Co., Inc.
- Toulmin, S.E., Rieke, R.D. & Janik, A. 1984. *An Introduction to Reasoning*. 2nd edn. London: Macmillan Publishing Co., Inc.
- Wertsch, J.V. 1985. *Culture, Communication, and Cognition: Vygotskian Perspectives*. Cambridge: Cambridge University Press.
- Wiliam, D. & Leahy, S. 2015. Embedding formative assessment: Practical techniques for K-12 classrooms. Learning Sciences International.
- Willis, J. & Klenowski, V. 2018. Classroom assessment practices and teacher learning: An Australian perspective. In Jiang, H. & Hill, M.F. (Eds.), *Teacher learning with classroom assessment: Perspectives from Asia Pacific*. London: Springer, 19-38.
- Xi, X. & Davis, L. 2016. Quality factors in language assessment. In Tsagari, D. & Banerjee, J. (Eds.), *Handbook of Second Language Assessment*. Munich: De Gruyter Mouton, 61-76.
- Xi, X. & Sawaki, Y. 2017. Methods of test validation. In Shohamy, E., Or, I.G. & May, S. (Eds.), *Language Testing and Assessment*. London: Springer International Publishing, 193-210.
- Yan, Z., Chiu, M.M. & Keung Cheng, E.C. 2022. Predicting teachers' formative assessment practices: Teacher personal and contextual factors. *Teaching and Teacher Education*, 114(2022): 103718.
- Zhao, D., Yan, B., Tang, L. & Zhou, Y. 2018. Teachers' understanding, implementing and reflecting upon classroom assessment in primary and junior high schools of China. In Jiang, H. & Hill, M.F. (Eds.), *Teacher learning with classroom assessment: Perspectives from Asia Pacific*. London: Springer, 77-98.